PROCEDURES FOR DEVELOPING A DEPTH-TO-GROUND WATER DATABASE

By

F. Spurlock

January 2000

STATE OF CALIFORNIA
Environmental Protection Agency
Department of Pesticide Regulation
Environmental Monitoring and Pest Management Branch
Environmental Hazards Assessment Program
Sacramento, California 95814-3510

EH 00-02

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

The Environmental Hazards Assessment Program of the Department of Pesticide Regulation (DPR) has developed statistical procedures for identifying soil and climatic characteristics of sections of land that are vulnerable to ground water contamination by pesticides (Troiano et al., 1998). Analysis of historical ground water monitoring data has also shown that depth to ground water (DGW, distance from the ground surface to first unconfined water, if any) is related to frequency of pesticide detection in ground water, especially for shallower DGW (Fig. 1). Consequently, EHAP's proposed CALVUL model to describe California's spatial vulnerability to ground water contamination includes DGW as one condition for determining spatial vulnerability. The CALVUL model has been used to focus EHAP's monitoring efforts (Troiano et al., 1999) and as the basis for identifying areas where modification of pesticide use practices will be required to mitigate ground water contamination by pesticides.

The depth-to-ground water (DGW) data in the EHAP DGW database are a relative measure of the depth-to-ground water at any given time. The relationship to vulnerability (as indicated by detection frequency) is based on the relative DGW of different sections. The current EHAP California DGW database (November 1999, Fig. 2) was developed from approximately 260,000 spring DGW measurements collected since 1987. These raw data were geostatistically analyzed to determine spatial autocorrelation so that optimum interpolation methods could be applied to generate sectional average DGW estimates. The data in the DGW database represent *relative* spring average sectional DGW estimates; they are not necessarily equal to actual measurements at any particular point in time and space due to the seasonal and annual variations in depth to ground water; however, the data should approximate the local depth to ground water in spring months.

Figure 1. Frequency of detection of 6800(a) compounds* in Fresno and Tulare Counties vs. depth to ground water

** simazine, atrazine, diuron, bromacil, triazine degradates and hexazinone.*

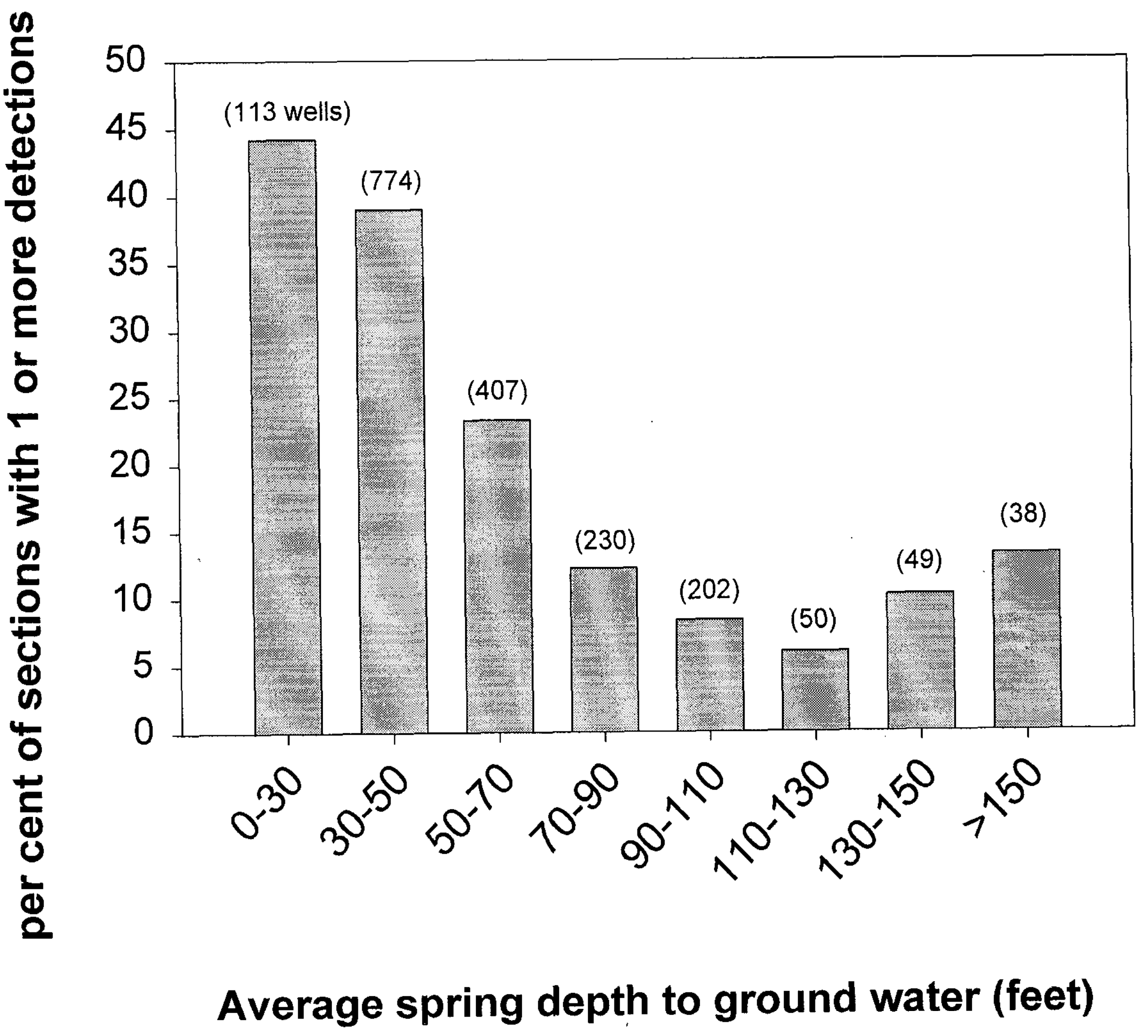
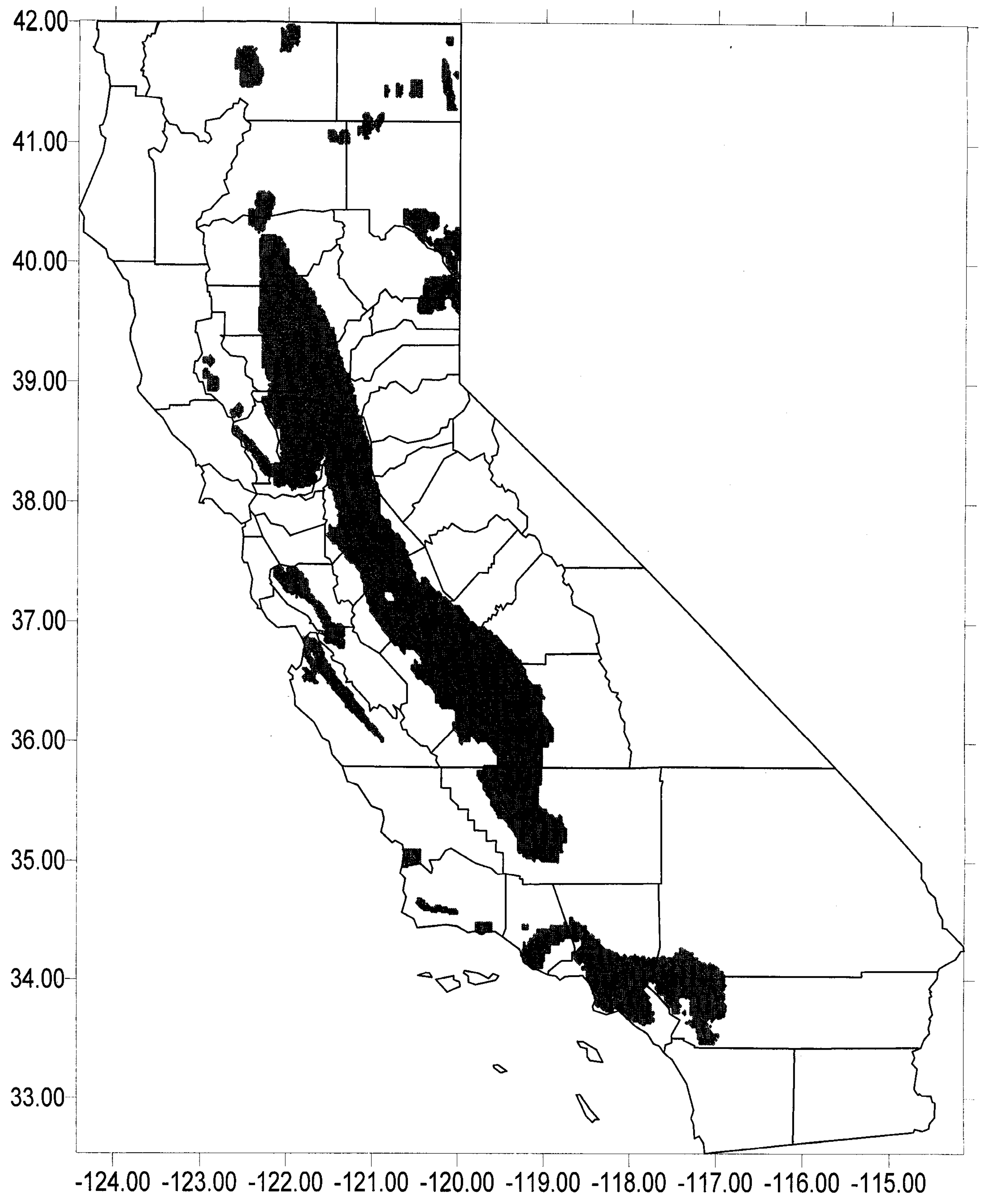


Figure 2. Post plot of sectional DGW coverage as of December 1999



The purpose of this report is to specify the procedures for

- ♦ obtaining raw DGW data,
- ♦ screening the data,
- ♦ conducting geostatistical analyses and gridding, and
- ♦ calculating sectional average spring DGW.

II. PROCEDURES

A. Sources of DGW data

A.1 The principal source of DGW data is the California Department of Water Resources, Division of Local Planning and Assistance (DWR DLPA). The four district offices and the counties for which they have data reported in the current EHAP DGW database are:

- ♦ Northern District (http://wwwdpla.water.ca.gov/nd)
  Butte, Colusa, Glenn, Lake, Lassen, Modoc, Plumas, Shasta, Siskiyou, Tehema

- ♦ Central District (http://wwwdpla.water.ca.gov/cd)
  Napa, Sacramento, San Joaquin, Solano, Sierra, Sutter, Yolo, Yuba

- ♦ San Joaquin District (http://wwwdpla.water.ca.gov/sjd)
  Fresno, Kern, Kings, Madera, Merced, Stanislaus, Tulare

- ♦ Southern District (http://wwwdpla.water.ca.gov/sd)
  Los Angeles, Orange, Riverside, San Bernadino, Santa Barbara, San Luis Obispo

Other sources of DGW data used to develop the current EHAP DGW database include the United States Geological Survey (U.S.G.S. 1995), San Benito County Water District, Santa Clara Valley Water District, and the Monterey County Water Resources Agency.

B. Minimum data requirements

B.1. When obtaining depth to ground water data, every effort should be made to obtain as much information as possible (e.g., appendix 1), including:

- state well number
- well latitude and longitude coordinates
- depth to ground water measured (or equivalent information such as reference point elevation, and measured distance from reference point to water surface)
- ground surface elevation
- basin code
- questionable measurement codes
- day/month/year of measurement
- agency making measurement

B.2. In some instances, not all information will be available. In these cases, the minimum required information is:

- state well number
- depth to ground water
- measurement date

C. Screening the data

C.1. Eliminate data records with questionable measurements (as given by questionable measurement codes, appendix 1).

C.2. Select measurement records from the months January through May.

C.3. Calculate decimal latitutde/longitude coordinates (ddlat, ddlong) for those wells with coordinate information given in degrees/minutes/seconds (° / ′ / ″) as follows:

decimal degrees = degrees + minutes/60 + seconds/3600

Note that the EHAP DGW database uses the convention that longitude coordinates are negative in California, increasing from west to east. Confirm that the longitude coordinates are negative in sign.

C.4. For well records without longitude/latitude coordinates, the coordinates may be approximated by the meridian/township/range/section/tract (MTRS-t) centroid coordinates. To determine the MTRS-t centroid coordinates:

C.4.a. Obtain the meridian/township/range/section (MTRS) centroid coordinates ($MTRS_{lat}$ and $MTRS_{long}$) from the oracle section centroid lat/long table (currently entitled *statepls*).

C.4.b. Calculate tract coordinate latitude and longitude adjustment factors (δ_{lat} , δ_{long}). These factors are determined as the local mean of the absolute value of:

$$\left| \frac{\Delta(ddlat)}{\Delta x} \right| = \delta_{lat}$$

$$\left| \frac{\Delta(ddlong)}{\Delta x} \right| = \delta_{long}$$

where x = distance (miles) and the respective differentials are taken in the direction of maximum change. The correction factors vary with location across the state, and are easiest to determine as the mean differences of ddlat and ddlong coordinates between adjacent sections of land in the vicinity of the well (e.g., from the township where the well is situated). As mentioned previously, MTRS centroid coordinates are available from the *statepls* database.

C.4.c. The MTRS-t centroid coordinates are then given by:

$$MTRS\text{-}t_{lat} = MTRS_{lat} + \boldsymbol{a}\,\delta_{lat}$$

$$MTRS\text{-}t_{long} = MTRS_{long} + \boldsymbol{b}\,\delta_{long}$$

where ***a*** and ***b*** are tabulated below according to tract.

Tract	Value of ***a***
A, B, C, D	+0.375
E, F, G, H	+0.125
J, K, L, M	-0.125
N, P, Q, R	-0.375

Tract	Value of ***b***
A, H, J, R	+0.375
B, G, K, Q	+0.125
C, F, L, P	-0.125
D, E, M, N	-0.375

C.5 Create a post plot of the data to verfy their areal coverage (e.g., Fig. 3).

If necessary, remove any data from isolated basins not hydrologically connected to the basin of interest.

Sources of information for identifying basins include:
basin codes (provided with most DWR data), DWR bull. 118 (1975), DWR bull. 118-80 (1980), U.S. Geological Survey maps, and common knowledge of isolating geographic structures such as mountain ranges.

C.6 Calculate means, and coefficient of variation (CV) of DGW for each well. Examine data with high CV to check for possible outlier values in the raw data.

C.7 Create a classed post plot of the data to examine major features of the data (Fig. 4). Observe any geographic patterns (or lack thereof) of the DGW data.

D. Geostatistical analyses

Three software programs with different geostatistical analytical methods/capabilities have been used to develop the current EHAP DGW database. These are: GEO-EAS (U.S. EPA, 1991, v. 1.2.1) , Surfer (Golden Software, 1996, v. 6.02), and GS+ (Gamma Design Software, 1999, v. 3.1a).

Figure 3. Post plot of observed depth-to-ground water data in San Joaquin Valley

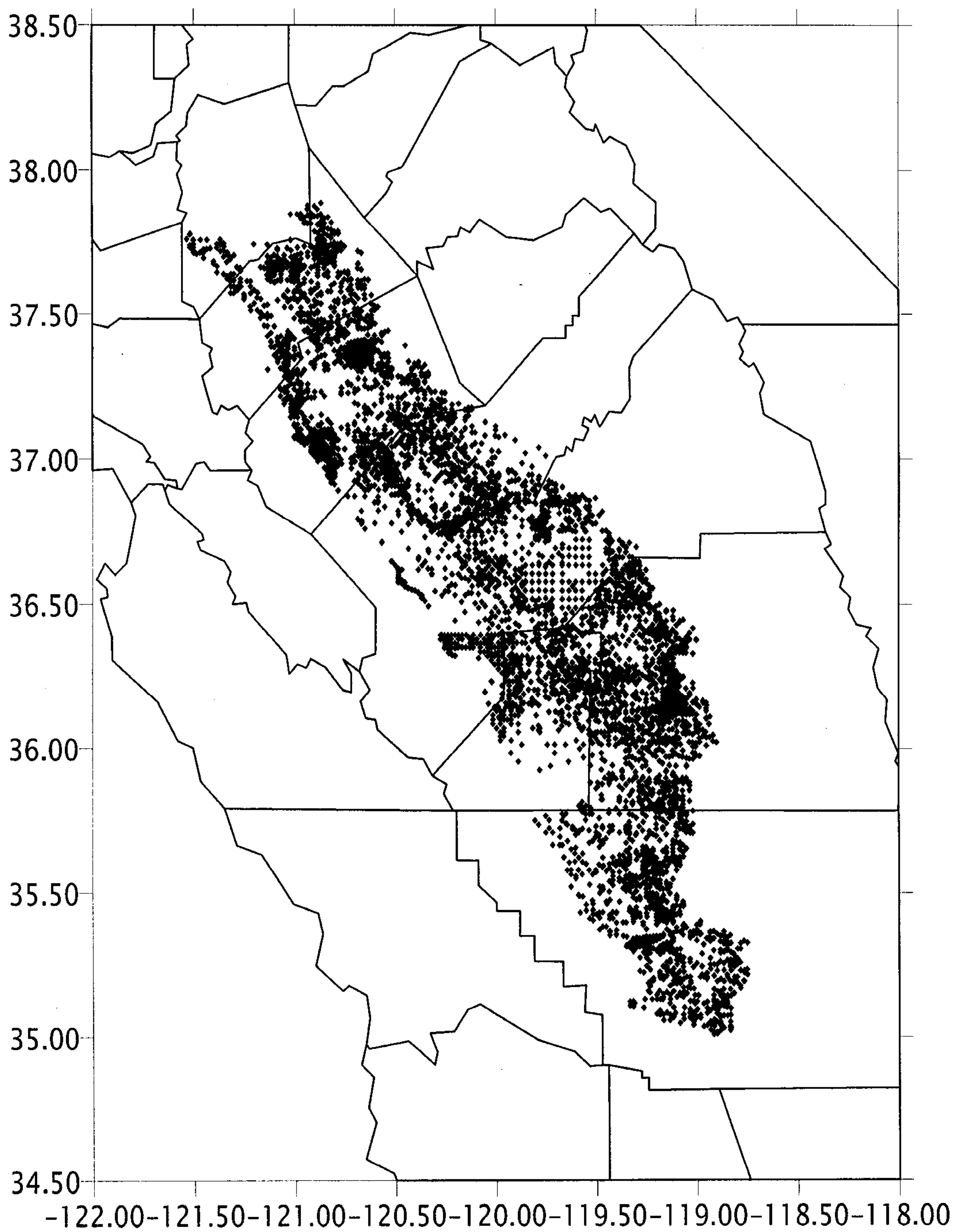
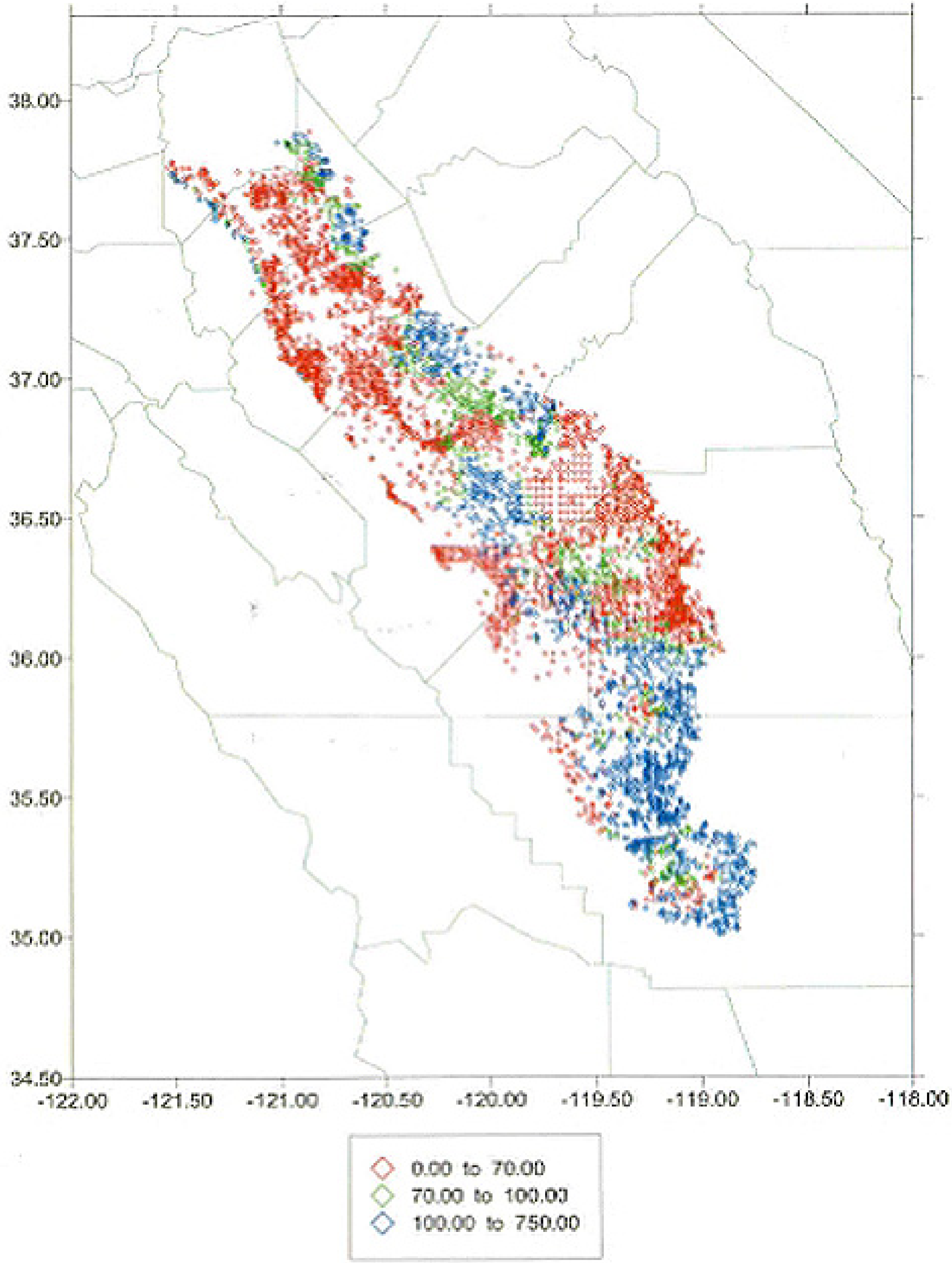


Figure 4. Classed post map of San Joaquin Valley depth to ground water data



The objective of the initial geostatistical analysis of the data is to identify the functional form and parameters of the semivariance vs. separation distance function. This function, called the variogram, is central to optimizing the interpolation of DGW from the measured data (Fig. 5). In addition, autocorrelograms may be developed to evaluate the dependence of autocorrelation on separation distance (Fig. 6). This information is useful for determining search ellipse radii for gridding the DGW data.

D.1. Calculate variograms and autocorrelograms to evaluate the spatial dependence of the data in preparation for gridding. Spatial dependence may depend on separation distance only (isotropic data) or may depend on both distance and direction (anisotropic data). Therefore, both isotropic and anisotropic variograms should be calculated to determine the extent of directional preference (if any) of autocorrelation (e.g., Fig. 5).

Also note: Variograms are often sensitive to the size of lag intervals; variograms of the data should be calculated using different lag intervals. The lag interval is the size of the distance "bin" that comprises the x-, or independent variable axis of the variogram plots. The total distance over which the variogram is plotted (the active lag distance) should fall in the general range of about 0.15 - 0.3 decimal degrees, which corresponds to approximately 10 – 20 miles. The variance structure beyond this general range is unimportant for estimating sectional average ground water depths.

E. Gridding

Kriging is generally considered to be a robust method of gridding, and is appropriate for most situations. To conduct kriging, select several different variogram models from those calculated above.

Figure 5. Isotropic and anisotropic variograms for Northern Sacramento DGW

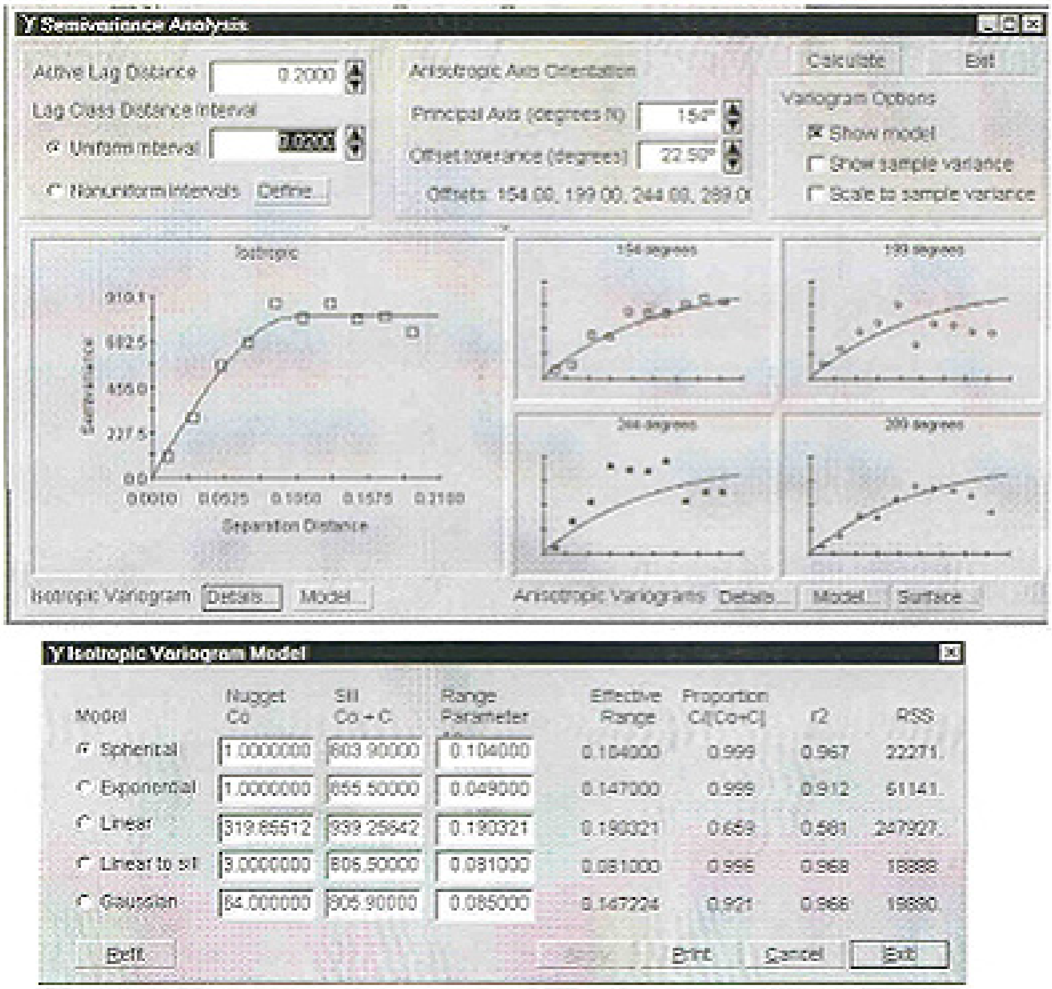
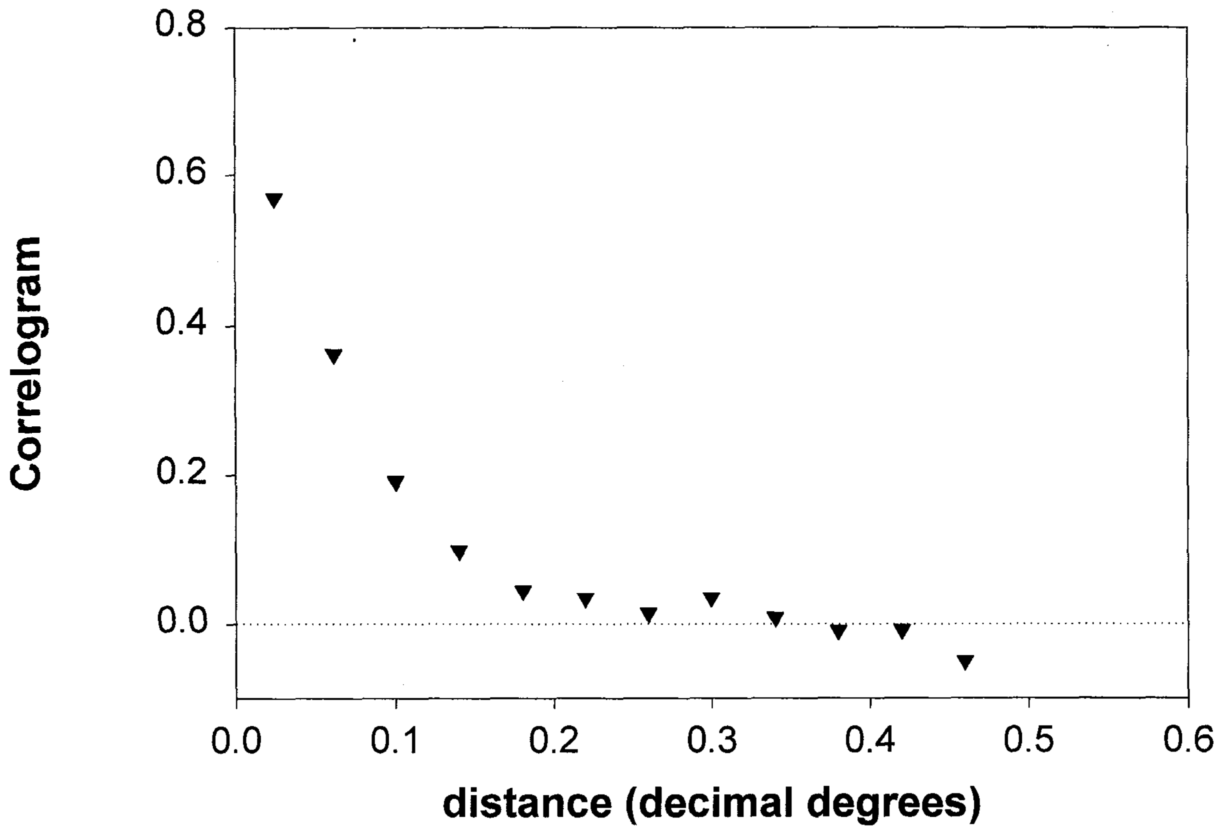


Figure 6. Correlogram for Sacramento Valley DGW data



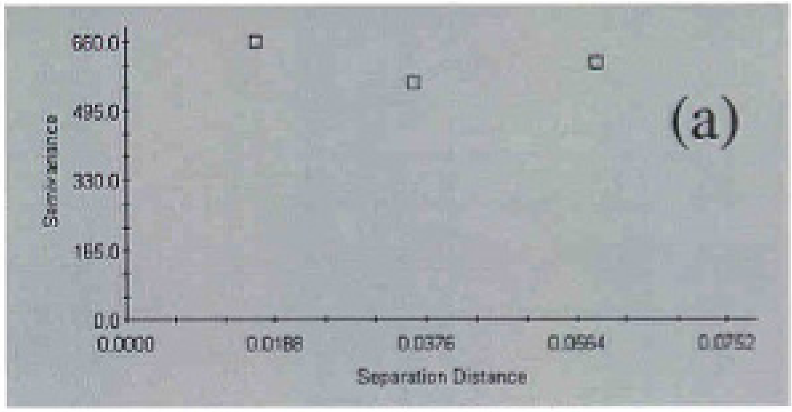
Note that:

- Spatial correlation of the data is not a requirement for gridding. In certain basins, all DGW data may be similar in magnitude in which case there will be little or no spatial autocorrelation (Fig. 7).
- Usually the data will be spatially autocorrelated; the variogram model chosen should accurately describe the variance structure over the separation distance range of the search ellipse radii; the ability of the variogram model to describe semivariance at large separation distances is unimportant.
- The search ellipse radius depends somewhat on the spatial density of the data as well as on data autocorrelation as indicated by the autocorrelogram. In general, smaller search radii are preferred; however spatial density of the data may also be a consideration. Some areas have a high density of DGW data, whereas other regions may have much lower data density. As a general guideline, typical radii values for DGW gridding range from about 0.03 to 0.10 decimal degrees (about 2 to 7 miles).

E.1 Grid the data.

E.2 Evaluate the different variogram models, anisotropy vs. isotropy, and search parameters by comparing sums of squared residuals (actual measured data – estimated data)^2 and the distribution of the residuals (Fig. 8) for the different griddings. A plot of observed vs. predicted DGW data (Fig. 9) along with regression line and prediction intervals is useful for visually determining potential outliers. The raw data for potential outliers should be examined.

E.3. Create a "classed" post plot of the residuals to examine the geographic distribution of the residuals for clustering or local grouping of large residual values (Fig. 10).



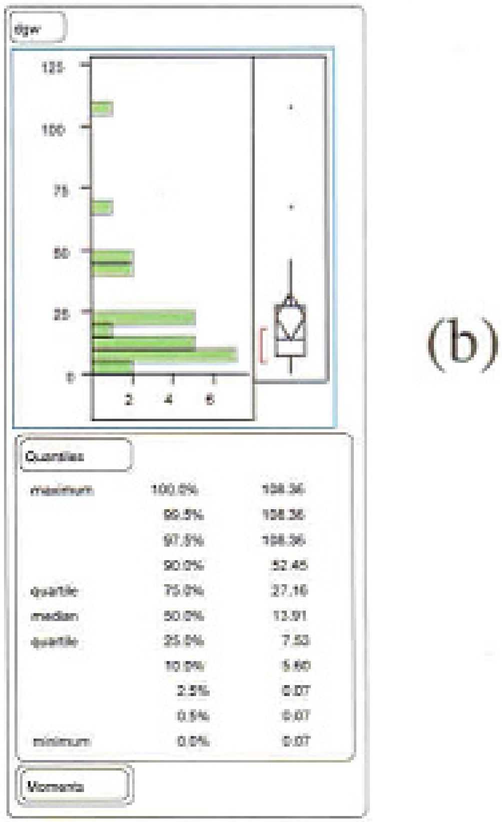


Figure 7. DGW data where (a) variogram indicates no spatial structure to the variance, (b) histogram shows nearly all data similar in magnitude, (c) classed post plot and gridded contours; most of the area is underlain by shallow ground water (<70 ft).

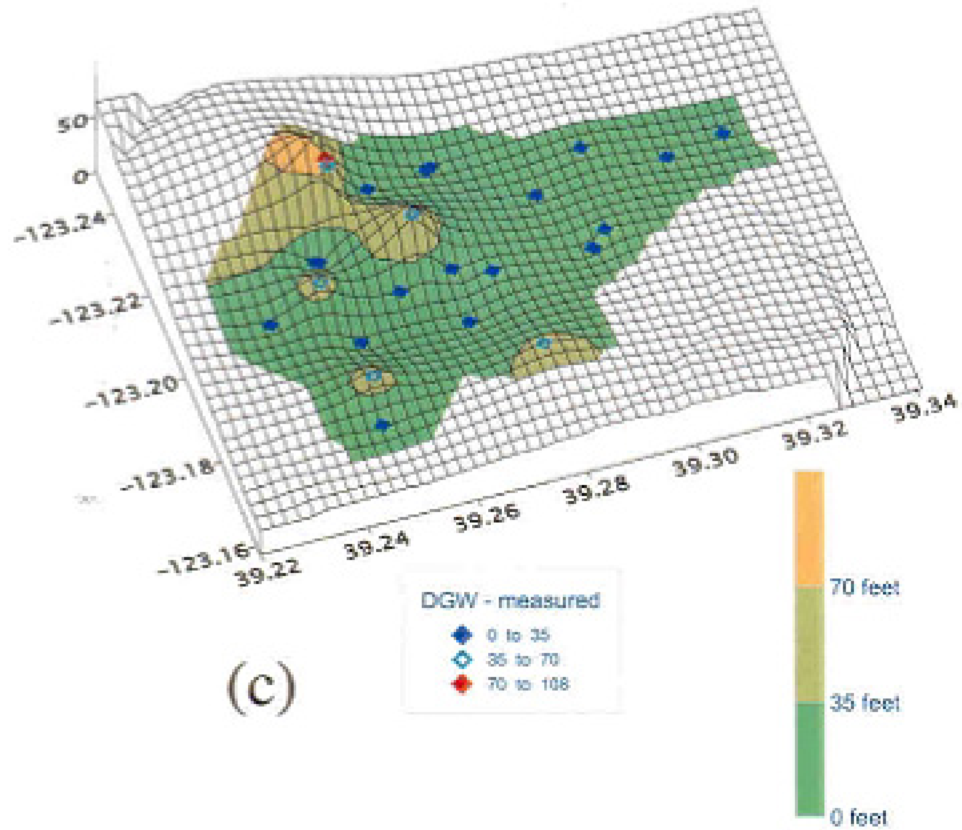
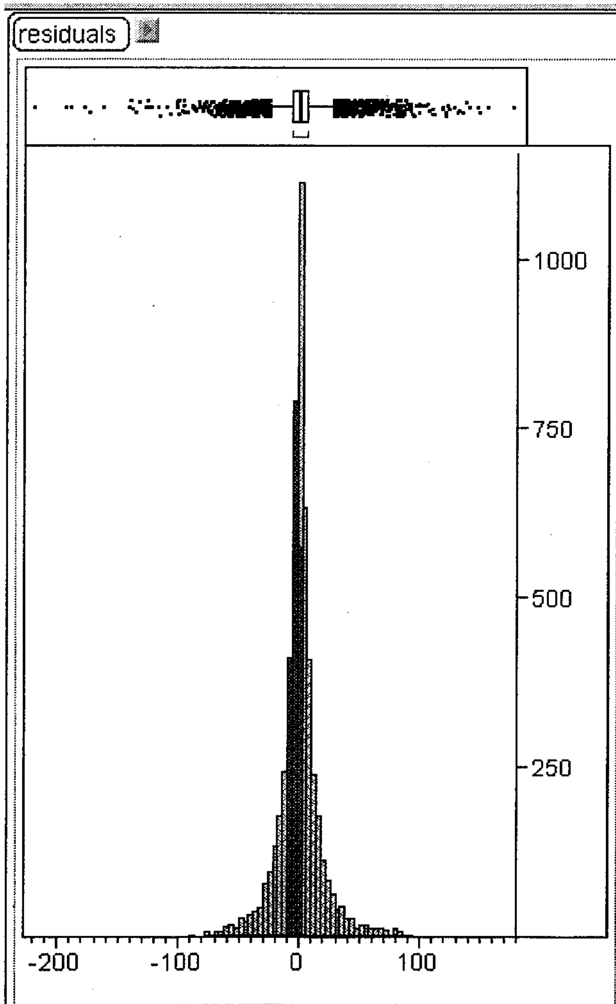


Figure 8. San Joaquin Valley DGW residuals (measured-gridded)



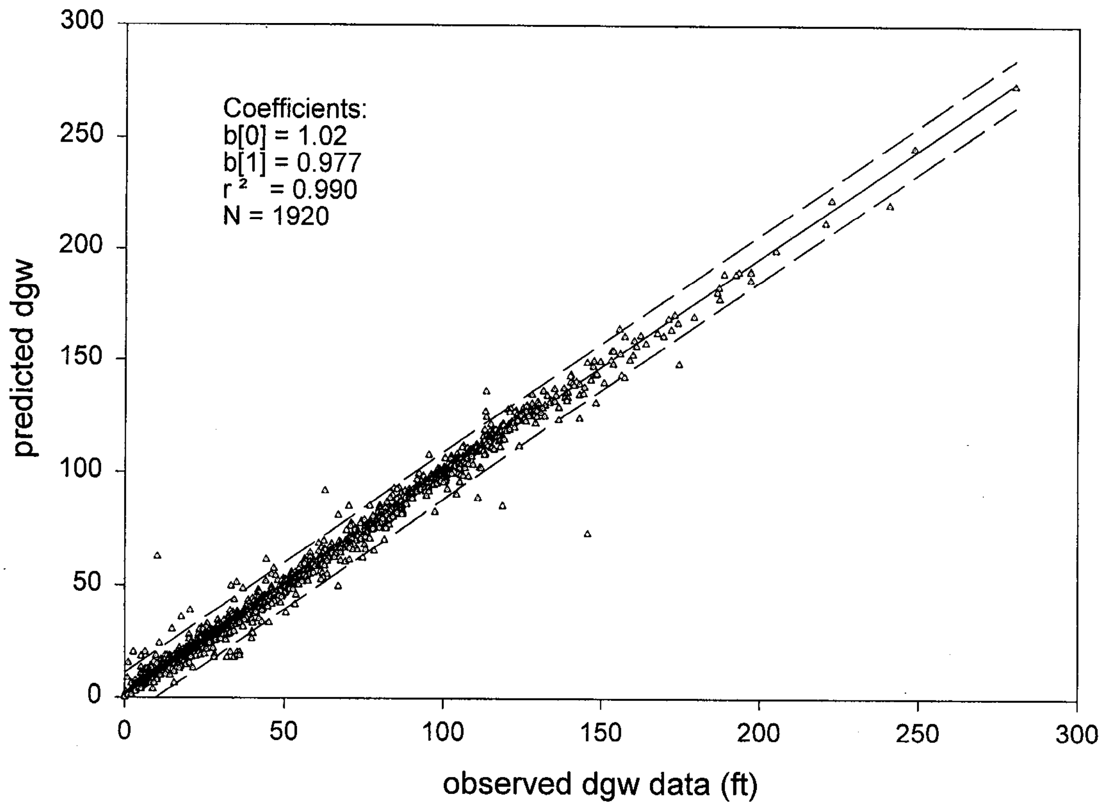
Quantiles

maximum	100.0%	176.78
	99.5%	97.79
	97.5%	57.43
	90.0%	20.35
quartile	75.0%	6.57
median	50.0%	-0.38
quartile	25.0%	-6.71
	10.0%	-20.35
	2.5%	-51.44
	0.5%	-96.74
minimum	0.0%	-218.72

Moments

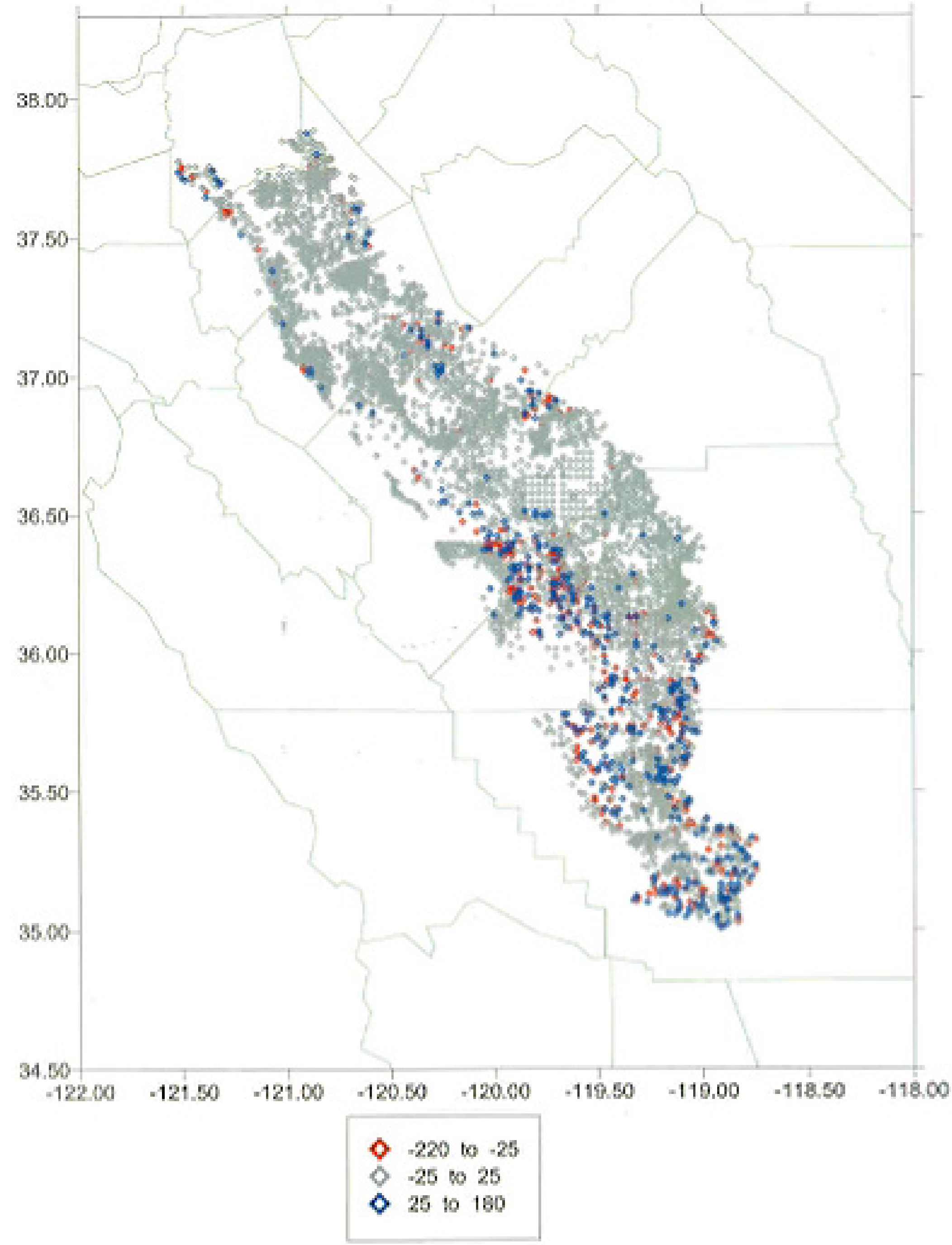
Mean	0.035
Std Dev	24.191
Std Error Mean	0.327
Upper 95% Mean	0.677
Lower 95% Mean	-0.606
N	5464.000
Sum Weights	5464.000

Figure 9. Comparison of observed and predicted northern Sacramento Valley DGW



		Stnd err	*student t*	*p*	*95% Confidence Intervals*	
Intercept	1.018	0.139	7.328	0.000	0.745	1.290
Slope	0.977	0.002	445.593	0.000	0.972	0.981

Figure 10. Classed post plot of residuals (measured - predicted) San Joaquin Valley DGW



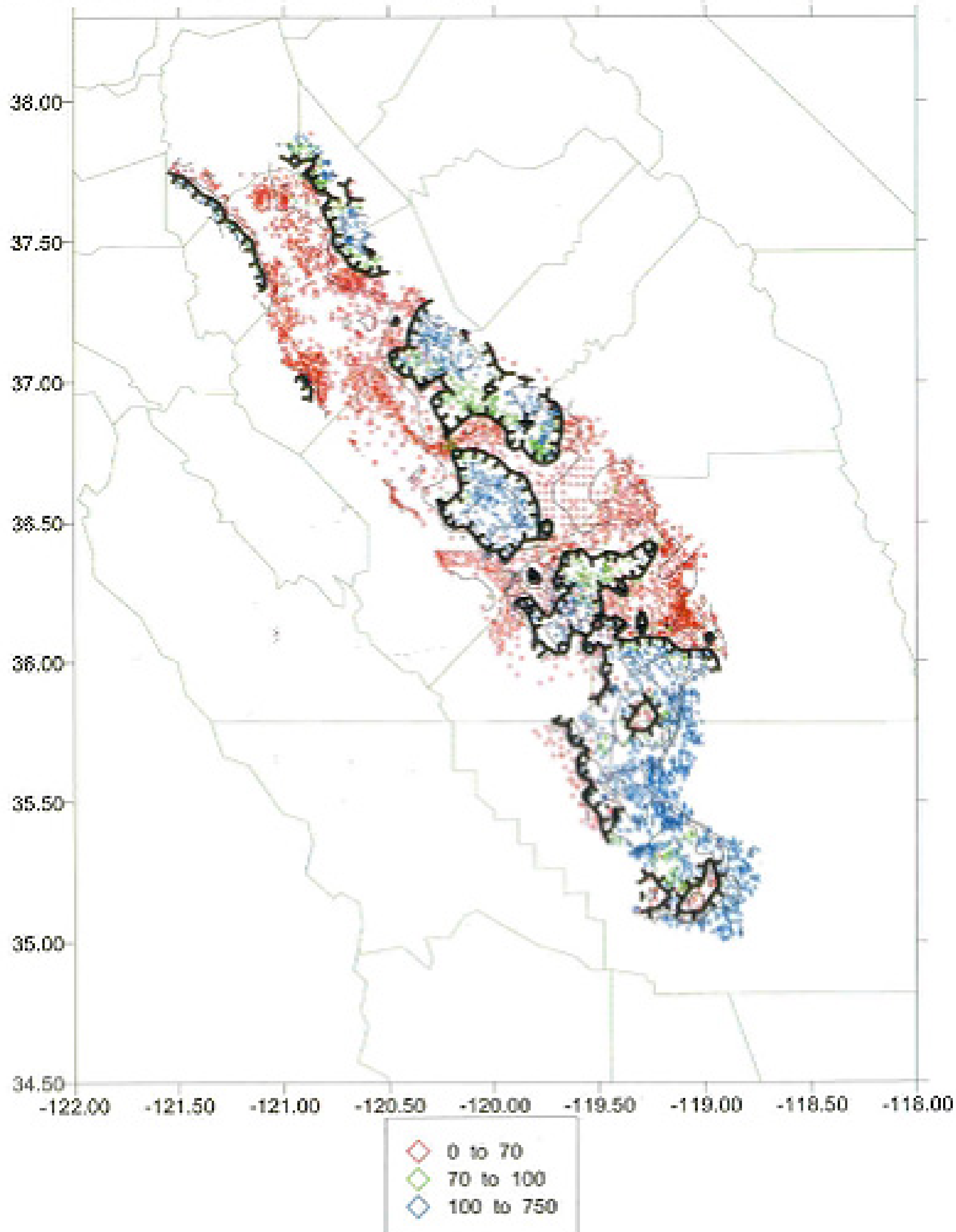
In addition, measured data associated with extreme residuals should be examined in the original raw data set to check for possible outliers.

E.4 An additional method for evaluating gridding output is to conduct cross-validation. The cross validation implementation in GEO-EAS is particularly useful providing the total number of data points does not exceed approximately 125. Further information is available in the GEO-EAS documentation.

F. Calculating sectional DGW averages

F.1 After the gridding is completed, calculate section centroid average spring DGW estimates using decimal longitude and decimal latitude coordinates and the gridding results. Create a contour map of the gridded values and superimpose a post plot of the raw DGW data to verify areal coverages of the raw and gridded data (Fig. 11); pay particular attention to borders where the gridding process may have determined estimated DGW values outside of the geographic range of the available raw data. As a general guideline, DGW estimates for areas that are greater than 1-2 miles outside the boundaries created by the raw data should be avoided.

Figure 11. Classed post plot of San Joaquin Valley DGW data and contours calculated from kriging. Hatched lines are 70 ft. contours.



REFERENCES

Troiano, J., C. Nordmark, T. Barry, and B. Johnson. 1997. Profiling areas of Ground Water Contamination by Pesticides in California: Phase II - Evaluation and Modification of a Statistical Model. Environ. Monitor. Assess. 45:301-318.

Troiano, J., J. Marade, and F. Spurlock. 1999. Empirical Modeling of spatial vulnerability applied to a norflurazon retrospective well study in California. J. of Environ. Qual. 28:397-403.

U.S. Geological Survey. 1995. Areal distribution of selected trace elements, salinity, and major ions in shallow ground water, Tulare Basin, sothern San Joaquin Valley, California. Water Resources Investigation publication 95-4048.

Dept. Water Resources. 1975. California's Ground Water. Bulletin 118.

Dept. Water Resources. 1980. Ground Water Basins in California. Bulletin 118-80.

Appendix 1. Description of Data in Ground Water Data Files

San Joaquin

DESCRIPTION OF DATA IN GROUND WATER DATA FILES

The Department of Water Resources sends out data files in ASCII delimited format. Each row contains one well measurement. Data columns are separated by commas. The first column contains the State Well Number. The second column contains the date in MM/DD/YYYY format that the well was measured. The third column contains the agency which measured the well. Agency codes are shown on the enclosed pages. The fourth column contains the No Measurement Code, if any, which is a reason why the well was not measured on this date. If there is no Code, this column is left blank. Explanations of No Measurement Codes are shown below. The fifth column contains the Questionable Measurement Code, if any, which is a reason why this well measurement may be questionable. If there is no Code, this column is left blank. Explanations of Questionable Measurement Codes are shown below. The sixth column contains the Ground to Water Surface Depth for this well. Measurements are in feet and are shown to the nearest tenth of a foot. The seventh column contains the Water Surface Elevation which is the depth in feet of the water surface above mean sea elevation. Measurements are also shown to the nearest tenth of a foot. The eighth column contains a single character text code which refers to the season this well was measured. "S" is for Spring, "F" is for Fall, and a blank is for any other season.

No Measurement Codes

0) Measurement discontinued	5) Unable to locate well
1) Pumping	6) Well has been destroyed
2) Pump house locked	7) Special
3) Tape hung up	8) Casing leaking or wet
4) Can't get tape in casing	9) Temporarily inaccessible

Questionable Measurement Codes

0) Caved or deepened	5) Air or pressure gage measurement
1) Pumping	6) Other
2) Nearby pump operating	7) Recharge operation at or nearby well
3) Casing leaking or wet	8) Oil in casing
4) Pumped recently	9) Acoustic sounder

Agency Code	Agency
1474	San Benito County
2855	Tenneco-West
3044	Tule River Association
5000	U.S. Geological Survey
5001	U.S. Bureau of Reclamation
5050	Department of Water Resources
5115	Monterey Co. Water Resources Agency
5128	Madera County
5129	Kings County Water District
5133	Kern County Water Agency
5200	City of Fresno
5203	City of Modesto
5282	Kern-Tulare Water District
5515	Central California Irrigation District
5520	Oakdale Irrigation District
5521	Modesto Irrigation District
5524	Turlock Irrigation District
5525	Merced Irrigation District
5525	Merced Irrigation District
5527	El Nido Irrigation District
5528	Chowchilla Water District
5529	Poso Resources Conservation District
5530	Madera Irrigation District
5531	San Luis Canal Company
5600	Orange Cove Irrigation District
5601	Stone Corral Irrigation District
5602	Ivanhoe Irrigation District
5603	Kaweah Delta Water Conservation District
5605	Exeter Irrigation District
5606	Lindsay-Strathmore Irrigation District
5607	Lindmore Irrigation District
5608	Porterville Irrigation District
5609	Lower Tule River Irrigation District
5611	Saucelito Irrigation District
5612	Pixley Irrigation District
5613	Delano-Earlimart Irrigation District
5614	So. San Joaquin Municipal Utility District
5615	North Kern Water Storage District
5616	Shafter-Wasco Irrigation District
5618	Corcoran Irrigation District
5619	Terra Bella Irrigation District
5620	James Irrigation District
5621	Tranquillity Resources Conservation District
5622	Garfield Water District
5623	Lewis Creek Water District
5626	Rag Gulch Water District
5631	Fresno Irrigation District
5636	Consolidated Irrigation District
5637	Alta Irrigation District
5640	Buena Vista Water Storage District
5644	Arvin-Edison Water Storage District
5646	Westlands Water District
5647	Tehachapi-Cummings County Water District
5649	Wheeler Ridge-Maricopa WSD
5701	California Water Service Company